



Páginas: 248-257
Recibido: 2021-10-06
Revisado: 2022-03-08
Aceptado: 2022-05-04
Publicación Final: 2022-09-15

www.revistascientificas.us.es/index.php/fuentes/index

DOI: <https://doi.org/10.12795/revistafuentes.2022.19232>

Adaptación y Validación de la Escala de Percepción del Clima de Evaluación en el Aula (EPCEA)

Adaptation and Validation of the Perceived Classroom Assessment Environment Scale (PCAES)

-   **Jorge Vergara-Morales**
Universidad de Las Américas (Chile)
-   **Claudio Díaz-Larenas**
Universidad de Concepción (Chile)
-   **Tania Tagle-Ochoa**
Universidad Católica de Temuco (Chile)
-   **Mabel Ortiz-Navarrete**
Universidad Católica de la Santísima Concepción (Chile)

Resumen

El clima de evaluación en el aula constituye un aspecto clave para analizar los aprendizajes. Para medir el constructo, actualmente se dispone de la Escala de Percepción del Clima de Evaluación en el Aula (EPCEA). Debido a que no se cuenta con antecedentes empíricos sobre su aplicación en el contexto educativo chileno, el objetivo del estudio es evaluar la estructura factorial, consistencia interna y validez de la EPCEA en estudiantes chilenos. Los participantes fueron 642 estudiantes de educación secundaria, 50.9% del género femenino y 49.1% del género masculino, con una edad media de 16.32 años. Se utilizó el modelo exploratorio de ecuaciones estructurales (ESEM) y el análisis factorial confirmatorio (AFC) para evaluar el modelo de medida. La consistencia interna se analizó mediante los coeficientes omega de McDonald y de fiabilidad compuesta. La validez convergente y discriminante se evaluaron mediante el cálculo de la varianza media extraída (VME) y de la \sqrt{VME} . Los resultados apoyaron la estructura de dos factores propuesta en el modelo original. No obstante, se re-especificó un modelo de medida de 14 ítems, que presentó adecuados niveles de fiabilidad y validez. Estos hallazgos sugieren que la adaptación de la versión en español de la EPCEA evidencia un modelo de medida con adecuadas propiedades psicométricas para medir la percepción sobre el clima de evaluación en el aula.

Abstract

The study of classroom assessment environments constitutes a key aspect of the learning process. To analyze this construct, the Perceived Classroom Assessment Environment Scale (PCAES) is currently available. Since there is no empirical evidence of the application of this instrument in the Chilean educational context, this study aims at assessing the factor structure, internal consistency, and validity of the PCAES among Chilean students. The participants were 642 high school students, 50.9% women and 49.1% men, with an average age of 16.32 years. The exploratory structural equation modeling (ESEM) and confirmatory factor analysis (CFA) were used to assess the fit of the measurement model to the observed data. Internal consistency analysis was carried out using both McDonald Omega and Composite Reliability coefficients. Convergent and discriminant validity were assessed by the measurement of Average Variance Extracted (VME) and \sqrt{VME} . Findings support the two-factor structure of the original model. However, a 14 item-model was re-specified that showed adequate levels of reliability and validity. These findings suggest that the Spanish adaptation of PCAES evidences a measure model of adequate psychometric properties to assess classroom assessment environment perceptions.

Palabras clave / Keywords

educación, clima de evaluación, evaluación del docente, educación secundaria, análisis multivariado, análisis factorial, validez interna, psicometría

education, classroom assessment environment, teacher evaluation, high school, multivariate analysis, factor analysis, internal validity, psychometry

1. Introducción

La evaluación del aprendizaje constituye un proceso continuo que involucra un conjunto variado de técnicas y procedimientos que permiten recoger evidencia de los desempeños estudiantiles. La evaluación es una parte integral del proceso de enseñanza y aprendizaje, que a lo largo del tiempo ha tenido la influencia de paradigmas como el conductista, centrado en la medición y utilización correcta y precisa de lo aprendido mediante técnicas y procedimientos que cuantifican y permiten observar las conductas de un estudiante (Gijbels & Dochy, 2006). Ha transitado también por el paradigma cognitivo, desde el cual se ha enfatizado el desarrollo de las habilidades y destrezas cognitivas necesarias para la ejecución de una tarea. Además, ha experimentado la influencia de perspectivas sociocognitivas, las que se han focalizado en el desarrollo de tareas y productos que involucran la colaboración y cooperación con los otros. En síntesis, se han constituido dos grandes macrocorrientes evaluativas (Mussawy, 2009).

Una de ellas, la llamada evaluación tradicional, enfatiza la prueba de papel y lápiz como una de las técnicas más efectivas para evaluar lo que un estudiante sabe. En esta lógica se encuentran todas aquellas técnicas evaluativas que conducen al estudiante a la demostración exenta de errores de lo que ha estudiado y aprendido secuencial y ordenadamente en clases. La segunda, denominada evaluación auténtica, busca evaluar los desempeños y productos de los estudiantes como evidencias de lo que son capaces de hacer o ejecutar (Muñoz & Alvares, 2007). La evaluación auténtica hace uso de todas aquellas técnicas evaluativas que permiten que el estudiantado demuestre desempeños similares a los que una persona realiza en la vida real. La evaluación auténtica, a diferencia de la evaluación tradicional, promueve la integración de saberes e involucra a los estudiantes en formas más complejas y profundas de aprendizaje.

Diferentes estudios muestran que el estudiantado debería tener un papel fundamental en el diseño del proceso evaluativo, puesto que ello los acerca y conecta con sus propios procesos de aprendizaje. Además, es importante considerar que muchas veces la evaluación genera ansiedad en los estudiantes, ya que frecuentemente se evidencia como una actividad desconectada de la enseñanza. Por esto, la evaluación como proceso debería seguir principios de validez, confiabilidad, transparencia, practicabilidad, autenticidad y retroalimentación. En otras palabras, se debe evaluar lo que se enseña y en la forma como se enseña, con instrucciones claras y transparentes que permitan que el estudiantado pueda responder adecuadamente a lo que se les está solicitando. En su diseño, la evaluación debería contener tareas auténticas similares a las del mundo real que, una vez evaluadas, puedan entregar una retroalimentación de calidad al estudiantado, considerando sus propios procesos de aprendizaje (Cheng et al., 2015).

El concepto de clima de evaluación fue introducido por Stiggins & Conklin (1992), y se define como el contexto de aula que experimenta el estudiantado cuando el profesorado determina las metas evaluativas, ejecuta las tareas de evaluación, define los criterios evaluativos, retroalimenta el aprendizaje y monitorea los resultados, es decir, es el clima de evaluación que percibe el estudiantado cuando el profesorado evalúa el aprendizaje (Alkharusi, 2011; Buldur & Doğan; 2014; Brookhart et al., 2006). El clima de evaluación puede afectar las percepciones de autoeficacia del estudiantado, sus metas de aprendizaje, y su desempeño académico. El clima evaluativo, según Alkharusi (2011) y Cheng et al. (2015), puede orientarse al aprendizaje o al desempeño.

Desde la perspectiva del estudiante, un clima de evaluación orientado al aprendizaje tiende a estar positivamente correlacionado con la autoeficacia y el logro académico, mientras que un clima de evaluación orientado al desempeño evidencia correlaciones negativas con dichas variables (Czura, 2017). A nivel docente, un clima de evaluación orientado al aprendizaje implica prácticas evaluativas que entreguen al estudiantado tareas significativas y de moderada dificultad, acompañadas de una rigurosa retroalimentación. Rescata las prácticas significativas de la evaluación del aprendizaje en el aula, con el fin de mejorar la construcción de conocimiento y el dominio de los contenidos. Un clima de evaluación orientado al desempeño, en cambio, conlleva al estudiante a una percepción de que las tareas evaluativas no están bien definidas, no son claras y tienen muy poco impacto en su vida estudiantil (Ali & Mohsin, 2013). Es decir, constituyen tareas que no tienen mucho que ver con los objetivos de aprendizaje de los estudiantes, por lo que no se sienten como partícipes del proceso evaluativo (Valva & Gokaj, 2013). El clima de evaluación orientado al desempeño se centra en las tareas evaluativas de mayor dificultad y menos significativas, que ponen el énfasis en las calificaciones sobre el aprendizaje, con el fin de comparar los desempeños de los estudiantes (Buldur & Doğan, 2014).

Desde una perspectiva motivacional, los planteamientos de la teoría de orientación a metas (Kaplan & Maehr, 2007) permiten apoyar que el clima de evaluación orientado al aprendizaje facilita la disposición de los estudiantes hacia el logro de las metas académicas. Es decir, permite la activación de comportamientos que

se orientan desde el propio control de la cognición, mediante la utilización de determinadas estrategias de aprendizaje. Esto provoca “niveles más altos de implicación cognitiva, afecto positivo e interés, más esfuerzo y persistencia en las tareas, búsqueda de ayuda y asunción de riesgos” (Alemán et al., 2011, p.160). Por otro lado, es posible plantear que el clima de evaluación orientado al desempeño promueve la disposición de los estudiantes hacia las metas de ejecución. Es decir, comportamientos centrados en demostrar habilidad mediante la comparación con los demás, o centrados en la evitación de evaluaciones negativas. Esto provoca “mayores niveles de ansiedad, menor valor de la tarea, menor implicación cognitiva, disminución del esfuerzo, dejar de persistir en la tarea, así como niveles más bajos de rendimiento” (Alemán et al., 2011, p. 160).

Al respecto, Ihan & Çetin (2014) exploraron la relación existente entre el clima de evaluación y *burnout* académico en 496 estudiantes de secundaria y descubrieron que, en climas de evaluación orientados al aprendizaje, el nivel de *burnout* académico disminuye y aumentan los niveles de logro académico. Por el contrario, en climas de evaluación orientados al desempeño el nivel de *burnout* académico aumenta y decrecen los niveles de logro académico. Buldur & Doğan (2014) describen la importancia del clima de evaluación en tanto este afecta la motivación, el sentido de logro y la autoeficacia del estudiantado. Es un concepto relevante en la literatura que no ha sido explorado en el contexto nacional, para el cual no existen instrumentos debidamente validados en la población chilena, particularmente si se considera que exámenes como el SIMCE, por ejemplo, tiene una lógica de evaluación por desempeño mediante una prueba nacional censal anual con altas consecuencias para las escuelas. La sigla SIMCE significa Sistema de Medición de la Calidad de la Educación, y fue establecido por ley en el año 1990 por el gobierno de Chile. Su propósito es evaluar a toda la cohorte de estudiantes de 4° y 8° grado de educación primaria y 2° grado de educación secundaria, de establecimientos privados, municipales o públicos y particular-subvencionado o privado-concertado. Cada año se evalúa el 4° grado de educación primaria en conjunto con uno de los demás grados, en las áreas de lenguaje y matemática. El área de evaluación en historia y ciencias naturales se considera exclusivamente para los grados 4° y 8° de educación primaria. “A partir de 2010 se incorporaron al sistema las áreas de Inglés como Segunda Lengua y Educación Física y se comenzó a evaluar el 2° y 6° grado de primaria” (Florez Petour, 2015, p. 32). Uno de los cuestionamientos al SIMCE “se refiere a los contenidos y preguntas de la prueba, los que no serían del todo relevantes, y modelarían negativamente la enseñanza” (Ortiz Cáceres, 2012, p. 357). Esto conduce a la necesidad de estudiar científicamente los efectos en el clima de evaluación que esto podría generar.

De acuerdo con la revisión de los antecedentes, se ha observado la utilización de la Escala de Percepción del Clima de Evaluación en el Aula (EPCEA) para evaluar la práctica evaluativa desde la perspectiva de los estudiantes. En la tabla 1, se presenta evidencia sobre la evaluación de las propiedades psicométricas de la EPCEA. Al respecto, se identifica que los diferentes trabajos se han centrado en analizar la estructura interna del instrumento, mediante la utilización del análisis factorial exploratorio. Por otro lado, la medición de la confiabilidad se ha centrado en la utilización del coeficiente alfa de Cronbach.

Tabla 1.
Evidencias psicométricas EPCEA

Autores	Tipo de análisis	Evidencias
Alkharusi (2011)	Análisis factorial exploratorio	Extracción por componentes principales con rotación varimax. La versión de 16 ítems evidenció una solución de dos factores que se ajustó adecuadamente a los datos observados. La solución factorial explicó entre 41.90% de la varianza total en las puntuaciones.
	Consistencia interna	Entre $\alpha = 0.75$; $\alpha = 0.82$
	Correlación entre factores	$r = -0.41$, $p < 0.01$
Alkharusi (2015)	Análisis factorial exploratorio	Extracción por componentes principales con rotación varimax. La versión original (16 ítems) evidenció una solución de dos factores que se ajustó adecuadamente a los datos observados. La solución factorial explicó entre el 19.68 (nivel del estudiante) y el 37.85% (nivel de clase) de la varianza total en las puntuaciones.

	Consistencia interna	Entre $\alpha = 0.65$ y $\alpha = 0.67$ (nivel del estudiante) Entre $\alpha = 0.70$ y $\alpha = 0.83$ (nivel de clase)
	Correlación entre factores	$r = -0.28$, $p < 0.01$ (nivel del estudiante) $r = -0.40$, $p < 0.01$ (nivel de clase)
Cheng et al., (2015)	Análisis factorial exploratorio	Extracción por componentes principales con rotación oblimin. La solución de dos factores de la versión original (16 ítems) se ajustó adecuadamente a los datos observados. La solución factorial explicó el 41% de la varianza total en las puntuaciones.
	Consistencia interna	Entre $\alpha = 0.70$ y $\alpha = 0.82$

El desarrollo inicial y validación de la EPCEA fue realizado por Alkharusi (2011), a través de un estudio que consideró una muestra de 450 estudiantes de escuelas públicas de Muscat, Omán. Mediante la revisión de la literatura científica sobre evaluación en el aula, se elaboraron 20 ítems asociados con tres aspectos de la evaluación que se relacionan directamente con la experiencia de los alumnos: (a) tareas de evaluación, (b) retroalimentación de la evaluación, y (c) estándares y criterios de evaluación. Los resultados iniciales mostraron que el modelo hipotético de tres dimensiones fue teóricamente menos significativo y difícil de interpretar. Luego de eliminar cuatro ítems de la escala, los resultados mostraron una solución de dos factores con niveles aceptables de consistencia interna. Los factores fueron denominados como: a) clima de evaluación orientado al aprendizaje, y b) clima de evaluación orientado al desempeño.

Posteriormente, Alkharusi (2015) realizó una nueva evaluación de las propiedades psicométricas de la EPCEA, considerando una muestra de 4088 estudiantes de 236 clases de segundo ciclo de las escuelas de educación básica de Omán. Si bien los resultados apoyaron la estructura interna de dos factores, algunos resultados de fiabilidad se situaron levemente por debajo del límite recomendado en la literatura ($\alpha = 0.70$). Por otro lado, Cheng et al. (2015) realizaron una adaptación de la EPCEA en una muestra de estudiantes universitarios de China. Los resultados mostraron que la solución de dos factores de la versión original se ajustó adecuadamente a los datos observados, mostrando niveles adecuados de consistencia interna.

A partir de los antecedentes revisados, se observan resultados convergentes sobre la validez de constructo de la EPCEA. No obstante, se observan diferencias en los valores de consistencia interna de las puntuaciones. Esto implica la necesidad de continuar obteniendo evidencia de las propiedades psicométricas de la EPCEA en diferentes contextos educacionales. Al situarse en el sistema educativo chileno, los antecedentes empíricos muestran que se carece de antecedentes sobre procesos de adaptación y validación de la EPCEA. Debido a que la percepción sobre el clima de evaluación en el aula tiene una marcada influencia en el proceso de aprendizaje, es importante disponer de instrumentos con propiedades psicométricas adecuadas para medir el constructo. De esta manera, el objetivo del estudio es evaluar la estructura factorial, fiabilidad y validez de la versión en español de la Escala de Percepción del Clima de Evaluación en el Aula (EPCEA) en estudiantes chilenos.

Respecto de los resultados del estudio, se espera disponer de evidencia estadística suficiente para apoyar la utilización de la EPCEA en el contexto educativo chileno, con el propósito de contribuir al mejoramiento de las prácticas evaluativas. De esta manera, resulta fundamental disponer de un instrumento con adecuadas propiedades psicométricas para evaluar el clima de evaluación que caracteriza el proceso de enseñanza y aprendizaje. Para el análisis, se consideraron los criterios de validez de Kane (2016), centrando el estudio en la evidencia de validez de estructura interna y de confiabilidad, con el propósito de realizar una interpretación válida sobre la percepción del clima de evaluación. En la línea de los planteamientos de Rios & Wells (2014), se establece que la orientación del estudio se ajusta a la naturaleza de los datos.

1.1. Hipótesis

En la Tabla 2, se muestran las hipótesis de la investigación centradas en la evaluación de la estructura interna y confiabilidad de la EPCEA. Para el caso del análisis exploratorio de ecuaciones estructurales y el análisis factorial confirmatorio, se espera el ajuste de un modelo de dos factores. Respecto de la evidencia de consistencia interna, se esperan valores de los coeficientes de omega de McDonald y de fiabilidad compuesta ≥ 0.70 . En relación con el análisis factorial confirmatorio multigrupo, se espera la invarianza del modelo de medida entre estudiantes de género femenino y masculino. Para la validez convergente de la estructura

interna se espera un valor de varianza media extraída (VME) > 0.40. Para el caso de la validez discriminante se espera un valor de $\sqrt{\text{VME}}$ superior al coeficiente de correlación entre los factores.

Tabla 2.
Evaluación de la estructura interna

Tipo de análisis	Evidencia esperada
Análisis exploratorio de ecuaciones estructurales (ESEM)	Modelo de dos factores: (1) clima de evaluación orientado al aprendizaje (2) clima de evaluación orientado al desempeño, con adecuados índices de ajuste a los datos observados.
Análisis factorial confirmatorio (AFC)	Modelo de dos factores: (1) clima de evaluación orientado al aprendizaje (2) clima de evaluación orientado al desempeño, con adecuados índices de ajuste a los datos observados.
Análisis factorial confirmatorio multigrupo (AFCM)	Invarianza del modelo de medida entre estudiantes de género femenino y masculino.
Consistencia interna	FC \geq 0.70 $\omega \geq$ 0.70
Validez de la estructura interna	VME > 0.40 $\sqrt{\text{VME}}$ es superior al coeficiente de correlación entre los factores del instrumento.

2. Metodología

2.1. Participantes

Los participantes del estudio fueron un total de 642 estudiantes de educación secundaria, quienes se encontraban cursando el 1° grado (29.1%), 2° grado (21.2%), 3° grado (3.7%) y 4° grado (46.0%). El 55% estudiaba en un establecimiento de tipo municipal o público (n = 352) y el 45% en un establecimiento de tipo particular-subsuccionado o privado-concertado (n = 290), localizados en la zona centro sur de Chile. El 50.9% de los estudiantes fueron del género femenino (n = 327) y el 49.1% fueron del género masculino (n = 315). La edad de los participantes se distribuyó entre los 13 y 20 años, con una edad media de 16.32 años (DE = 1.59). Se utilizó un muestreo no probabilístico de tipo incidental, considerando cohortes intactas para la selección de los participantes.

2.2. Instrumento

Escala de Percepción del Clima de Evaluación en el Aula (EPCEA). Este instrumento mide la percepción de los estudiantes sobre el clima de evaluación en el aula. Se compone de 16 ítems que se distribuyen en dos factores: (a) clima de evaluación orientado al aprendizaje (9 ítems, p.ej.: “los estudiantes tenemos la oportunidad de corregir nuestros errores”), (b) clima de evaluación orientado al desempeño (7 ítems, p.ej.: “El profesor(a) da más importancia a las notas que al aprendizaje”). Los ítems se responden en una escala Likert de 5 puntos (1 = Muy en desacuerdo; 5 = Muy de acuerdo). Las puntuaciones del instrumento han presentado una adecuada fiabilidad para ambos factores (Alkharusi, 2011). Los ítems fueron contextualizados para la medición en el ámbito de la asignatura de inglés mediante la consigna “Durante la clase de inglés...”. Los ítems fueron respondidos considerando una escala Likert de 5 puntos (1= Muy en desacuerdo; 5= Muy de acuerdo).

2.3. Procedimiento

La EPCEA fue adaptada al idioma español mediante el uso de la metodología de traducción directa e inversa (traducción-retrotraducción). Para su desarrollo, se consideraron las siguientes etapas (Beaton et al., 2000):

- (a) Traducción al español: dos traductores profesionales de lengua materna española, y bilingües en inglés, tradujeron de forma independiente la versión original en inglés de la escala.
- (b) Retro-traducción al idioma original: un traductor de lengua materna inglesa y bilingüe en español realizó una retrotraducción o traducción inversa. Se compararon la versión retrotraducida con la versión original, para detectar diferencias o discrepancias entre ambas y proponer nuevas formulaciones de los ítems en español.

La versión en español de la EPCEA fue aplicada en la sala de clases de los estudiantes, previa autorización declarada por las autoridades escolares correspondientes. El instrumento fue aplicado por un profesional previamente capacitado, considerando la participación voluntaria de los estudiantes. La investigación se realizó siguiendo los criterios éticos de la American Psychological Association (APA), lo que se materializó en la firma del consentimiento informado. Durante la aplicación del instrumento, se informó a los estudiantes sobre el estricto resguardo de la confidencialidad de sus respuestas, estipulándose su exclusiva utilización para fines académicos. La investigación se orientó desde un diseño instrumental, ya que se buscó analizar las propiedades psicométricas del modelo de medida de la EPCEA en su versión en español. De acuerdo al proceso de recolección de los datos, se obtuvo una tasa de respuesta del 100% debido a que la aplicación del instrumento se basó en cohortes intactas. Finalmente, la investigación fue aprobada por el Comité de Ética, Bioética y Bioseguridad de la Universidad de Concepción, Chile.

2.4. Análisis de los datos

Se realizó un análisis descriptivo mediante el cálculo de la media, la desviación estándar, las medidas de asimetría y curtosis, además del porcentaje de datos perdidos. La consistencia interna de los datos se evaluó mediante el coeficiente Omega de McDonald (ω) y de Fiabilidad Compuesta (FC), considerando un límite inferior de 0.70 para identificar una fiabilidad aceptable (Hair et al., 2010). La estructura factorial de la EPCEA se evaluó mediante estrategia de validación cruzada, dividiendo aleatoriamente la muestra en dos mitades de igual tamaño. La primera sub-muestra ($n = 321$) se consideró para realizar un análisis exploratorio de ecuaciones estructurales (ESEM), utilizando el método de estimación de máxima verosimilitud (ML – Muthén & Muthén, 2012). Además, se utilizó el método de rotación oblicua (geomin), debido a que se consideró la correlación entre los factores (Costello & Osborne, 2005). Se consideraron las cargas factoriales estandarizadas > 0.40 (Abós Catalán et al., 2018). La segunda sub-muestra ($n = 321$) se consideró para realizar un análisis factorial confirmatorio (AFC), utilizando el método de estimación ML para evaluar el ajuste del modelo hipotético a los datos observados.

Tanto para el ESEM como para el AFC, la adecuación del modelo factorial se evaluó considerando los siguientes índices y criterios: (a) χ^2/df : un buen ajuste se indica con valores inferiores a 3; (b) Índice de Ajuste Comparativo (CFI) e índice de Tucker-Lewis (TLI): un ajuste aceptable se indica a través de valores ≥ 0.90 , y un buen ajuste está determinado por valores ≥ 0.95 ; (c) Error Cuadrático Medio de Aproximación (RMSEA): un ajuste aceptable está determinado por valores ≤ 0.08 (90% IC ≤ 0.10), y un buen ajuste se indica mediante valores ≤ 0.06 (90% IC ≤ 0.08) (Kelloway, 2015).

La equivalencia del modelo de medida de la EPCEA se evaluó mediante análisis factorial confirmatorio multi-grupo (AFCM) (Brown, 2006), para lo cual se consideró la totalidad de los participantes ($N = 642$). El AFCM se realizó mediante una evaluación secuencial de la invarianza configural, métrica y escalar. La invarianza del modelo de medida es aceptada si los valores de CFI presentan una variación ≤ 0.01 en relación al modelo anterior (Cheung & Rensvold, 2002). Los análisis fueron realizados utilizando el programa estadístico Mplus versión 8.

La validez de la estructura interna se evaluó a través de la validez convergente y discriminante. La validez convergente se analizó a través del cálculo de la varianza media extraída (VME), medida que muestra la relación entre la varianza que es capturada por un factor determinado respecto de la varianza total debida al error de medida de ese factor (Cheung & Wang, 2017). Un valor VME > 0.40 permite aceptar la validez convergente (Aldás, 2013). La validez discriminante se analizó mediante el cálculo de la \sqrt{VME} , medida que

evalúa si un factor determinado difiere de otros constructos. Se acepta la validez discriminante si la \sqrt{VME} es superior al coeficiente de correlación entre los factores (Henseler et al., 2015).

3. Resultados

3.1. Análisis descriptivo

En la tabla 3 se observa que las puntuaciones medias muestran una tendencia favorable hacia la percepción de un clima de evaluación orientado al aprendizaje ($M = 3.73$). Además, se observa que las puntuaciones medias muestran adecuados niveles de dispersión, ya que los valores de desviación estándar (DE) fueron cercanos a 1. Por otro lado, las medidas de asimetría y curtosis permiten inferir una tendencia hacia la normalidad univariada, ya que los valores obtenidos fueron inferiores a ± 2 (George & Mallery, 2010). Finalmente, los datos perdidos (DP) se ubicaron entre el 2% y 3%, los cuales fueron reemplazados mediante el método imputación bayesiano. Estos resultados apoyan la utilización del método de estimación de máxima verosimilitud para los análisis factoriales exploratorios y confirmatorios.

Tabla 3.
Estadísticos descriptivos

Factores	M	DE	Asimetría	Curtosis	% DP
1. Clima de evaluación orientado al aprendizaje	3.73	0.70	-0.70	0.72	3%
2. Clima de evaluación orientado al desempeño	2.62	0.74	0.22	-0.02	2%

3.2. Análisis exploratorio de ecuaciones estructurales

Los resultados de la evaluación del modelo original de dos factores (16 ítems) mostraron un adecuado nivel de ajuste a los datos observados ($\chi^2/df = 1.70$; CFI= 0.94; TLI= 0.92; RMSEA= 0.05). No obstante, se identificaron cargas factoriales menores a 0.40. Debido a esto, se decidió re-especificar el modelo de medida removiendo los ítems 7 y 9. Los resultados del modelo re-especificado permiten aceptar el ajuste del modelo de medida a los datos observados, ya que los valores de los índices estuvieron dentro de los límites recomendados ($\chi^2/df = 2.00$; CFI= 0.94; TLI= 0.91; RMSEA= 0.06). Por otro lado, se observó que las variables se representaron adecuadamente en el espacio bi-factorial, ya que todas las cargas factoriales fueron > 0.40 (ver tabla 4).

Tabla 4.
Distribución de cargas factoriales según modelos testeados

Ítems	Modelo 2 factores (16 ítems)		Modelo 2 factores (14 ítems)	
	F1	F2	F1	F2
1	0.60*	-0.04	0.61*	-0.02
2	0.59*	0.02	0.61*	0.04
3	0.70*	-0.04	0.70*	-0.04
4	0.74*	0.01	0.74*	0.01
5	0.62*	-0.03	0.62*	-0.02
6	0.55*	0.03	0.55*	0.04
7	0.27*	0.24	-	-
8	0.56*	0.00	0.56*	0.01
9	0.36*	0.06	-	-
10	0.01	0.53*	0.01	0.53*
11	0.16	0.50*	0.16	0.48*
12	-0.04	0.49*	-0.04	0.49*
13	0.06	0.63*	0.07	0.62*
14	-0.19	0.44*	-0.18	0.46*
15	-0.23	0.58*	-0.22	0.58*
16	-0.07	0.43*	-0.04	0.46*

*p < 0.05

3.2. Análisis factorial confirmatorio

Los resultados de la evaluación del modelo original de dos factores (16 ítems) no permiten aceptar el ajuste del modelo de medida a los datos observados, ya que se observan valores que se sitúan fuera de los límites recomendados (CFI = 0.897 y TLI = 0.880). Debido a esto, se decidió re-especificar el modelo de medida, removiendo uno a la vez los ítems con cargas factoriales < 0.40. Para cada caso, se verificó la significancia estadística y el ajuste del modelo de medida a los datos observados. De acuerdo con los resultados del proceso de re-especificación, se obtuvo un modelo de medida de 14 ítems con aceptables niveles de ajuste a los datos observados ($\chi^2/df = 2.20$; CFI= 0.93; TLI= 0.92; RMSEA= 0.06)

Los ítems del modelo re-especificado estuvieron fuertemente correlacionados con las variables latentes que miden, ya que las cargas factoriales presentaron valores > 0.40. Además, fueron estadísticamente significativas a un nivel $p < 0.01$ (ver figura 1).

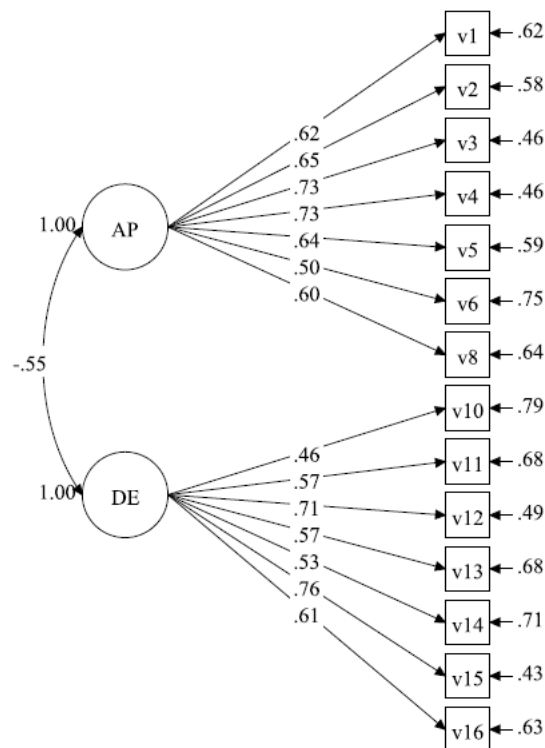


Figura 1. Coeficientes estandarizados modelo re-especificado (14 ítems)

Nota. AP = Clima de evaluación orientado al aprendizaje; DE = Clima de evaluación orientado al desempeño

3.3. Análisis factorial confirmatorio multigrupo

En la tabla 5 se muestran los resultados del análisis de invarianza del modelo de medida entre estudiantes de género masculino y femenino. Para el análisis, se consideraron los siguientes aspectos: (1) un modelo base que establece el mismo patrón de carga factorial para los dos grupos (invarianza configural); (2) se restringió el modelo base sobre las cargas factoriales evaluándose la invarianza métrica; y (3) se incorporó una restricción sobre los interceptos de los ítems para evaluar la invarianza escalar. De acuerdo con los valores de los índices de ajuste, se acepta la invarianza del modelo de medida entre los estudiantes de género femenino y masculino, debido a que la variación en los valores CFI no fue superior a 0.01. Además, los valores de χ^2/gf , TLI y RMSEA se sitúan dentro de los límites recomendados.

Tabla 5.
Medidas de invarianza por género

Modelos de invarianza	χ^2/gl	CFI	ΔCFI	TLI
Configural	1.80	0.95	-	0.94
Métrica	1.90	0.94	0.01	0.93
Escalar	1.95	0.93	0.01	0.93

3.4. Análisis de fiabilidad y validez

Los resultados del análisis de fiabilidad muestran que el modelo de dos factores re-especificado presentó puntuaciones con adecuados niveles de consistencia interna, ya que los valores de los coeficientes de fiabilidad compuesta (FC) y de omega de McDonald (ω) fueron > 0.70 . Por otro lado, para el caso del factor clima de evaluación orientado al aprendizaje, se observa que el valor de VME fue > 0.40 , por lo que se acepta la validez convergente. Para el caso del factor clima de evaluación orientado al desempeño, se identifica que el valor de VME es levemente inferior a 0.40. No obstante, debido a que los valores de consistencia interna se sitúan en 0.80, es posible aceptar la validez convergente (Moral-de la Rubia, 2019). Finalmente, se identifica que los valores de la \sqrt{VME} fueron superiores al coeficiente de correlación entre los factores de la EPCEA ($r = 0.40$), por lo que se acepta la validez discriminante del modelo de medida (ver tabla 6).

Tabla 6.
Correlaciones, fiabilidad y validez

Factores	1	2	ω	FC	VME	\sqrt{VME}
1. Clima de evaluación orientado al aprendizaje	1.00		0.83	0.83	0.41	0.64
2. Clima de evaluación orientado al desempeño	-0.40**	1.00	0.80	0.80	0.37	0.61

Nota. ** $p < 0,01$

4. Discusión

El objetivo del estudio fue evaluar la estructura factorial, fiabilidad y validez de la versión en español de la Escala de Percepción del Clima de Evaluación en el Aula (EPCEA) en estudiantes chilenos. La investigación surgió a partir de la necesidad de ampliar la disponibilidad de la EPCEA para medir la percepción de los estudiantes sobre el clima de evaluación en el contexto educativo chileno. Los resultados del ESEM y del AFC mostraron que la solución de dos factores se ajustó adecuadamente a los datos observados. No obstante, fue necesario re-especificar el modelo de medida a través de la eliminación de ítems. El modelo re-especificado presentó adecuados niveles de ajuste a los datos observados, mostrando puntuaciones factoriales con apropiados niveles de fiabilidad y validez. Estos resultados coinciden con los hallazgos de Alkharusi (2011), ya que la composición de la EPCEA se redujo en el número de ítems. Si bien no se pudo replicar por completo la estructura factorial original, se conservó la estructura del modelo de medida de dos factores, lo que coincide con los hallazgos de Alkharusi (2011) y Cheng et al. (2015). De acuerdo con la revisión de los antecedentes, se puede plantear que las dificultades de ajuste del modelo de medida podrían estar ocasionadas por la existencia de sesgos en las respuestas al instrumento, provocadas por las diferencias tanto culturales como relacionadas con el clima de evaluación. Al respecto, los resultados apoyan parcialmente la primera y segunda hipótesis de investigación, ya que la evaluación de la estructura interna de la EPCEA evidenció una estructura factorial bastante similar a la versión original. Por otro lado, los resultados apoyan la confirmación de la tercera hipótesis, ya que las puntuaciones del modelo de medida re-especificado de la ECEA presentaron niveles adecuados de consistencia interna. Finalmente, los resultados confirman la cuarta hipótesis de investigación, ya que el modelo de medida re-especificado evidenció valores adecuados de validez convergente y discriminante. Por lo tanto, los resultados de evaluación de la estructura interna y confiabilidad de la EPCEA han sido satisfactorios, ya que se determinó una estructura factorial bastante similar a la versión original, con adecuados niveles de fiabilidad.

5. Conclusiones

A partir del estudio, se concluye que los hallazgos apoyan un modelo de medida re-especificado de la EPCEA, que permite obtener mediciones fiables y válidas sobre la percepción del clima de evaluación en estudiantes chilenos de educación secundaria. Los hallazgos del estudio constituyen un aporte para el proceso de evaluación del aprendizaje, ya que permiten valorar el contexto de aula que experimenta el estudiantado cuando el profesorado determina las metas evaluativas, ejecuta las tareas de evaluación, define los criterios evaluativos, retroalimenta el aprendizaje y monitorea los resultados (Alkharusi, 2011; Buldur & Doğan; 2014). De esta manera, la identificación del tipo de clima de evaluación que prevalece en las aulas chilenas facilitará la indagación de los procesos cognitivos, motivacionales y emocionales que influyen en la evaluación de los aprendizajes. Esto permitirá promover el mejoramiento de las prácticas evaluativas, lo que afectará positivamente las percepciones de autoeficacia del estudiantado, sus metas de aprendizaje, y su desempeño académico, ya que se facilitará la articulación del clima de evaluación con los propios procesos de aprendizaje de los estudiantes (Cheng et al., 2015).

Respecto de las limitaciones, se considera que las puntuaciones de la EPCEA no fueron correlacionadas con variables externas. De esta manera, se sugiere considerar instrumentos de medida como el Inventario de Percepción de las Actividades de Evaluación [PATI - Perception of Assessment Tasks Inventory] (Dorman & Knightley, 2006) o el Inventario de Preferencias de Evaluación [API - Assessment Preference Inventory] (Birenbaum, 1994), ya que permitirá aportar mayor evidencia de validez del modelo de medida. Se considera importante que futuras investigaciones consideren la influencia de la clase en la percepción del clima de evaluación en el aula, por lo que se recomienda la utilización del análisis factorial confirmatorio multinivel del instrumento de medida. Finalmente, se recomienda la utilización de la EPCEA considerando las puntuaciones de los ítems de cada factor, ya que constituyen medidas de distintos tipos de clima de evaluación en el aula. Además, se recomienda su aplicación con estudiantes de enseñanza media de establecimientos municipales y/o particulares-subvencionados, con el fin de identificar el tipo de clima de evaluación que perciben los estudiantes sobre la base de las metas evaluativas, las tareas de evaluación, los criterios evaluativos, el monitoreo y la retroalimentación de los resultados que realizan los docentes.

Apoyos

Este estudio se inserta en el contexto del proyecto FONDECYT 1191021, *Estudio correlacional y propuesta de intervención en evaluación del aprendizaje del inglés: las dimensiones cognitiva, afectiva y social del proceso evaluativo del idioma extranjero*.

Referencias

- Abos Catalán, A., Sevil Serrano, J., Martín-Albo Lucas, J., Julián Clemente, J. A., & García-González, L. (2018). An integrative framework to validate the Need-Supportive Teaching Style Scale (NSTSS) in secondary teachers through exploratory structural equation modeling. *Contemporary Educational Psychology*, 52, 48-60. <https://doi.org/10.1016/j.cedpsych.2018.01.001>
- Aldás, J. (2013). La invarianza del instrumento de medida. In F. Sarabia (Ed.), *Metodología para la investigación en marketing y dirección de empresas* (pp. 386-410). Pirámide.
- Alemán, M. J., Trías, D., & Curione, K. (2011). Orientaciones motivacionales, rendimiento académico y género en estudiantes de bachillerato. *Ciencias Psicológicas*, 5(2), 159-166. <http://www.scielo.edu.uy/pdf/cp/v5n2/v5n2a04.pdf>
- Ali, M. & Mohsin, M. (2013). Test anxiety inventory (TAI): Factor analysis and psychometric properties. *Journal of Humanities and Social Science*, 8(1), 73-81. <https://www.iosrjournals.org/iosr-jhss/papers/Vol8-issue1/K0817381.pdf>
- Alkharusi, H. (2011). Development and datametric properties of a Scale measuring students' perceptions of the Classroom assessment environment. *International Journal of Instruction*, 4(1), 105-120. <https://eric.ed.gov/?id=ED522908>
- Alkharusi, H. (2015). An Evaluation of the Measurement of Perceived Classroom Assessment Environment. *International Journal of Instruction*, 8(2), 45-54. <https://doi.org/10.12973/iji.2015.824a>
- Beaton D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191. <https://doi.org/10.1097/00007632-200012150-00014>
- Birenbaum, M. (1994). Toward adaptive assessment—the student's angle. *Studies in Educational Evaluation*, 20, 239-255. [https://doi.org/10.1016/0191-491X\(94\)90011-6](https://doi.org/10.1016/0191-491X(94)90011-6)

- Brookhart, S. M., Walsh, J. M., & Zientarski, W. A. (2006). The dynamics of motivation and effort for classroom assessments in middle school science and social studies. *Applied Measurement in Education*, 19(2), 151–184. https://doi.org/10.1207/s15324818ame1902_5
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Buldur, S. & Doğan, A. (2014). Adaptation of the students' perceptions of the Science and Technology course classroom assessment environment scale into Turkish. *Education & Science/Eğitim ve Bilim*, 39(176), 199-211. <https://doi.org/10.15390/EB.2014.3729>
- Cheng, L., Wu, Y., & Liu, X. (2015). Chinese university students' perceptions of assessment tasks and classroom assessment environment. *Language Testing in Asia*, 5 (1). <https://doi.org/10.1186/s40468-015-0020-6>
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, & Evaluation*, 10, 1-9. <https://doi.org/10.7275/yj1-4868>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Cheung, G. W., & Wang, C. (2017). Current Approaches for Assessing Convergent and Discriminant Validity with SEM. *Issues and Solutions. Proceedings*, 2017(1). <https://doi.org/10.5465/AMBPP.2017.12706abstract>
- Czura, A. (2017). Adolescent learner perceptions of foreign language assessment: Critical incident analysis. *Glottodidactica. An International Journal of Applied Linguistics*, 44(2), 25-39. <https://pressto.amu.edu.pl/index.php/gl/article/view/12092/11939>
- Dorman, J. P., & Knightley, W. M. (2006). Development and validation of an instrument to assess secondary school students' perceptions of assessment tasks. *Educational Studies*, 32(1), 47–58. <https://doi.org/10.1080/03055690500415951>
- Flórez Petour, M. T. (2015). Validity and equity in educational measurement: The case of SIMCE. *Psicoperspectivas*, 14(3), 31-44. <https://doi.org/10.5027/PSICOPERSPECTIVAS-VOL14-ISSUE3-FULLTEXT-618>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equations models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50. <https://doi.org/10.2307/3151312>
- George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference, 17.0 update*. Allyn & Bacon
- Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: can formative assessment make a difference?. *Educational studies*, 32(4), 399-409. <https://doi.org/10.1080/03055690600850354>
- Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2010). *Multivariate Data Analysis* (7th ed.). Pearson Prentice Hall.
- Henseler, J., Ringle, C.M. & Sarstedt, M. A. (2015). New criterion for assessing discriminant validity in variance-based structural equation modeling. *J. of the Acad. Mark. Sci.*, 43, 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Ihan, M., & Çetin, B. (2014). An analysis of the relationship between academic burnout and classroom assessment environment. *Ted Eğitim Ve Bilim*, 39 (176), 51-68.
- Kane, M. T. (2016). Explicating validity, Assessment in Education. *Principles, Policy & Practice*, 23(2), 198-211. <https://doi.org/10.1080/0969594X.2015.1060192>
- Kaplan, A., & Maehr, M. L. (2007). The Contributions and Prospects of Goal Orientation Theory. *Educational Psychology Review*, 19(2), 141-184.
- Kelloway, E. K. (2015). *Using Mplus for structural equation modeling*. SAGE Publications, Inc. <https://doi.org/10.4135/9781483381664>
- Kline, R. (2016). *Principles and practice of structural equation modeling* (4^a ed.). The Guilford Press.
- Moral-de la Rubia, J. (2019). Revisión de los criterios para validez convergente estimada a través de la Varianza Media Extraída. *Psychologia*, 13(2), 25-41. <https://doi.org/10.21500/19002386.4119>
- Muñoz, A., & Alvares, M. (2007). Students' objectivity and perception of self-assessment in an EFL classroom. *The Journal of Asia TEFL*, 4(2), 1-25.
- Mussawy, S. (2009). *Assessment practices: Students' and teachers' perceptions of classroom assessment* (Unpublished Master's Thesis). School of Education, University of Massachusetts.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (7th ed.). Muthén & Muthén.
- Ortiz Cáceres, I. (2012). En torno a la validez del Sistema de Medición de la Calidad de la Educación en Chile. *Estudios pedagógicos (Valdivia)*, 38(2), 355-373. <https://doi.org/10.4067/S0718-07052012000200022>
- Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 108-116. <https://doi.org/10.7334/psicothema2013.260>
- Stiggins, R. & Conklin, N. (1992). *Investigating the practices of classroom assessment*. State University of New York Press.
- Vavla, L., & Gokaj, R. (2013). Learner's perceptions of assessment and testing in EFL classrooms in Albania. *Mediterranean Journal of Social Sciences*, 4(11), 509. <https://doi.org/10.5901/mjss.2013.v4n11p509>